

PENGENALAN CITRA DOKUMEN SASTRA JAWA **(Studi Kasus Pada Buku Sastra Jawa: Menak** **Sorangan I dan Panji Sekar)**

Document Image Recognition Of Javanese Manuscripts
(Case study in Javanese Manuscripts: Menak Sorangan I and
Panji Sekar)

Anastasia Rita Widiarti¹ dan Agus Harjoko²

Program Studi Ilmu Komputer
Sekolah Pascasarjana Universitas Gadjah Mada

ABSTRACT

Yogyakarta hosts many collections of classic Javanese manuscripts that are invaluable cultural heritage. Converting these texts into digital format will give immense benefits, particularly in terms of text enhancement and preservation.

This research aimed to make a prototype of document image recognition system for Javanese manuscripts, which consisted of three steps. First, a given texts were read using optic devices, and saved as picture-type files *.jpg, which were then fed into the system as input data. The second step is to perform pixel-level processing on the input data to prepare them for further analysis. Such processing includes: binarization to reduce a gray scale to a binary image; orientation normalization to correct the skew of the document which typically results from improper paper feeding into the scanner; filling to reduce noise; thinning to get the skeleton of each character; and segmentation to separate document image into its components down to individual characters. The third step is character recognition to derive the meaning of the characters. Character recognition algorithms have two essential components: (i) feature extraction, i.e. by counting the number of pixels of the object in each unit of a character image, the characteristics of the image were determined. These

¹ Fakultas MIPA Universitas Sanata Darma, Yogyakarta

² Fakultas MIPA Universitas Gadjah Mada, Yogyakarta

characteristics were kept in the Javanese character database, which afterwards was used in the verification of a given character, and (ii) classification, i.e. by utilizing the modification of the Euclidean distance to match the Javanese character with its corresponding Latin character.

The research showed that the success rate of this system in recognizing the images of the given Javanese manuscripts, in this case Menak Sorangan I and Panji Sekar pages 3 and 4, was 81.28%. In conclusion, it can be said that the methods used in the stages of this document image recognition system has been appropriate.

Keywords : *Character recognition. Document image recognition. Pixel-level processing.*

PENGANTAR

Yogyakarta adalah daerah yang terkenal akan kekayaan sejarah budayanya, termasuk kesenian, dan kesusastaan. Salah satu kekayaan sejarah yang tidak ternilai harganya adalah naskah-naskah Jawa yang masih banyak dijumpai di Yogyakarta, yaitu di Kraton Kasultanan, dan Pura Pakualaman (Suryakusuma, 2003). Apabila naskah-naskah tersebut dapat dikonversikan ke dalam format digital akan banyak manfaat yang dapat diraih, antara lain kualitasnya dapat diperbaiki, keberadaannya dapat diperpanjang, serta dapat dimanipulasi untuk kepentingan penelitian, misalnya untuk diterjemahkan secara otomatis ke dalam tulisan latin tanpa mengubah makna naskah asli.

Perkembangan ilmu analisis citra dokumen, yaitu analisis pada representasi visual dokumen kertas seperti jurnal, hasil faksimili, surat-surat kantor, lembar isian, dan lain-lain (Srihari, *et al.*, 1986), membuka peluang besar untuk dimanfaatkan bagi pelestarian naskah-naskah kuno yang banyak ditemukan di Yogyakarta. O'Gorman dan Kasturi (1997) memberikan tahapan-tahapan proses analisis citra dokumen yang dapat dimodifikasi untuk pengenalan citra dokumen sastra Jawa. Dimulai dengan tahap pengambilan di mana data dari dokumen kertas akan dibaca dengan alat scan optik dan hasilnya disimpan sebagai *file* citra. Kemudian dilanjutkan dengan tahap pengolahan tingkat piksel

yang bertujuan untuk menyiapkan dokumen citra, serta membuat fitur perantara untuk membantu mengenali citra. Tahap yang ketiga adalah tahap pengenalan karakter dengan tujuan untuk menerjemahkan sederetan karakter yang memiliki berbagai macam bentuk dan ukuran.

Sejauh yang penulis ketahui, penelitian yang berkait dengan aplikasi pengenalan citra dokumen sudah banyak dilakukan. Sauvola *et al.* (1997) menyodorkan sebuah metode baru untuk binarisasi citra dokumen secara adaptif. Kavallieratou *et al.* (2005) mengaplikasikan distribusi kelas Cohen's untuk menaksir besar sudut kemiringan suatu citra dokumen. Azar (1997) membahas penggunaan algoritma Hilditch untuk menemukan bentuk dasar suatu karakter Latin. Flether dan Kasturi (1988) mengembangkan algoritma Robust untuk memisahkan teks dari suatu citra yang berisi campuran teks dan gambar. O'Gorman (1993) mengaplikasikan sebuah metode untuk melakukan analisis tata letak halaman pada artikel-artikel PAMI. Brown (2000) mengaplikasikan algoritma *Feature Point Extraction* untuk pengenalan karakter Latin. Suprihatin (2003), dalam tesisnya telah berhasil mengalihaksarakan tulisan karakter Jawa ke tulisan karakter Latin, namun masukan untuk program alihaksara diperoleh dari *keyboard*.

Pada tulisan ini disajikan salah satu cara rekayasa perangkat lunak pada citra dokumen teks Jawa modern, meliputi detail tahap pengolahan tingkat piksel pada teks Jawa modern, kemudian dilanjutkan dengan tahap ekstraksi ciri karakter Jawa, dan tahap terakhir yaitu tahap pencocokan ciri karakter Jawa.

CARA PENELITIAN

Tahapan penelitian yang digunakan dalam penelitian ini mempergunakan modifikasi proses analisis citra dokumen yang dikembangkan oleh O'Gorman dan Kasturi (1997). Persiapan yang dilakukan sebelum pelaksanaan penelitian berupa studi literatur, pendalaman MATLAB, serta menentukan masalah yang akan dianalisis. Data yang dipergunakan dalam pengujian program adalah data dari dokumen buku sastra Jawa Menak Sorangan I (Yasadipura I, 1936) dan Panji Sekar halaman 3 dan 4 (Pakubuwono IV, 1933)(lihat Tabel 1.).

Tabel 1. Data citra dokumen hasil pembacaan alat scan optis

Asal Dokumen	Dimensi Matriks Citra Dokumen (baris vs kolom)	Besar File (byte)
Menak Sorangan Halaman 3	2208 vs 1502	493056
Menak Sorangan Halaman 4	2312 vs 1520	3514240
Panji Sekar Halaman 3	2323 vs 1496	3475208
Panji Sekar Halaman 4	2307 vs 1505	3472035

Pengujian program dilakukan untuk mengetahui kebenaran hasil pengenalan citra dokumen. Hasil keluaran yang dihasilkan program diteliti dan dibandingkan dengan hasil pengenalan secara manual oleh seorang ahli sastra Jawa, sehingga dapat diketahui apakah program yang dibuat sudah benar dan cukup memadai.

Sebagai pembandingan hasil keluaran program aplikasi MATLAB, dilakukan perhitungan pada hasil pembacaan data secara manual oleh ahli sastra Jawa pada data dan kondisi kualitas cetakan data yang sama. Hasil keluaran program serta pembacaan manual ditampilkan dalam bentuk tabel. Untuk mengetahui apakah hasil keluaran program serta hasil pembacaan manual mempunyai hasil yang sama, dilakukan penelitian satu persatu terhadap setiap perbedaan hasil pengenalan karakter dari dokumen.

Tahap pengenalan citra dokumen diawali dengan menangkap atau mengambil citra dokumen teks Jawa dengan mempergunakan scanner. Hasil pengambilan citra tersebut kemudian disimpan dalam file gambar monokrom bertipe *.jpg. File gambar ini kemudian akan menjadi masukan dalam program pengenalan citra dokumen untuk diproses lebih lanjut.

Tahapan berikutnya adalah tahap pengujian program, yang akan dimulai dengan proses pengolahan tingkat piksel. Pada tahap ini akan dilakukan proses binarisasi, kemudian normalisasi orientasi, *filling*, perangkaan, dan terakhir adalah proses segmentasi.

Proses binarisasi bertujuan untuk memisahkan citra utama atau citra obyek dari citra latar belakang yang tidak dibutuhkan (O'Gorman dan Kasturi, 1997). Secara umum persamaan yang dipakai untuk menghasilkan citra biner adalah sebagai berikut.

$$g(x,y) = \begin{cases} 1 & \text{jika } f(x,y) > k \\ 0 & \text{jika } f(x,y) \leq k \end{cases} \quad (1)$$

dengan $g(x,y)$ adalah citra biner dari citra monokrom $f(x,y)$, dan k menyatakan nilai ambang. Nilai ambang k ditentukan dengan menggunakan metode otsu (Otsu, 2003) karena metode ini dapat menentukan secara otomatis nilai ambang berdasarkan citra masukan. Nilai ambang k akan berkisar antara 1 sampai dengan L , dengan $L = 255$, dapat ditentukan dengan memaksimumkan persamaan:

$$\sigma_B^2(k^*) = \max_{1 \leq k \leq L} \sigma_B^2(k) \quad (2)$$

$$\text{dengan: } \sigma_B^2(k) = \frac{[\mu_T \omega(k) - \mu(k)]^2}{\omega(k)[1 - \omega(k)]} \quad (3)$$

$$\mu_T = \sum_{i=1}^L i \cdot p_i, \quad \mu(k) = \sum_{i=1}^k i \cdot p_i, \quad \omega(k) = \sum_{i=1}^k p_i, \quad p_i = \frac{n_i}{N} \quad (4)$$

di mana n_i menyatakan jumlah piksel dengan tingkat keabuan i , dan N menyatakan banyaknya piksel pada area citra.

Proses normalisasi orientasi bertujuan untuk mengurangi pengaruh kemiringan posisi dokumen saat pengambilan gambar. Dimulai dengan menentukan sudut kemiringan θ dari citra biner dengan mempergunakan persamaan (5) sebagai berikut (Jain, 1989):

$$\theta = \frac{1}{2} \tan^{-1} \left[\frac{2\mu_{1,1}}{\mu_{2,0} - \mu_{0,2}} \right] \quad (5)$$

$$\text{dengan: } \mu_{p,q} = \sum_m \sum_n (m - \bar{m})^p (n - \bar{n})^q \quad (6)$$

merupakan momen pusat ordo (p,q) , dan \bar{m} , \bar{n} merupakan titik pusat momen yang didefinisikan sebagai berikut:

$$\bar{m} = \frac{1}{N} \sum_m \sum_n m, \quad \bar{n} = \frac{1}{N} \sum_m \sum_n n, \quad (7)$$

di mana N menyatakan jumlah piksel pada citra, m dan n menyatakan koordinat obyek dari citra. Kemudian citra biner diputar sebesar θ dengan mempergunakan persamaan (8):

$$\begin{aligned} x' &= x \cos(\theta) - y \sin(\theta) \\ y' &= x \sin(\theta) + y \cos(\theta) \end{aligned} \quad (8)$$

Hasil dari proses normalisasi orientasi ini adalah citra biner yang ternormalisasi.

Tahap selanjutnya dalam pengolahan piksel adalah tahap untuk mengurangi *noise* atau disebut juga tahap *filling*. Dua hal pokok yang dilakukan pada tahap ini adalah proses dilatasi dan erosi, di mana kedua proses ini mendasarkan pada 8 titik tetangga dari suatu piksel. Erosi adalah pengurangan ukuran dari daerah ON, yaitu sebagai akibat adanya piksel-piksel ON yang berada di daerah OFF, dan dilatasi adalah proses sebaliknya dari erosi, yaitu menambahkan piksel ON untuk menutup daerah ON (O'Gorman dan Kasturi, 1997).

Setelah dilakukan proses *filling*, pengujian dilanjutkan dengan menjalankan proses perangkaan. Perangkaan atau *thinning* atau istilah lainnya *medial-axis* adalah suatu proses pengurangan komponen-komponen citra dengan tujuan untuk mendapatkan informasi yang paling mendasar tentang pembentuk citra atau untuk mendapatkan kerangka suatu citra. Dalam penelitian ini algoritma *thinning* yang dipergunakan adalah algoritma Hilditch (Azar, 1997).

Tahap terakhir dalam pengolahan tingkat piksel adalah tahap segmentasi. Tahap ini bertujuan untuk memperoleh citra-citra karakter Jawa yang terdapat dalam citra dokumen masukan. Pada penelitian ini pendekatan yang dipakai untuk mendapatkan citra-citra karakter adalah dengan mempergunakan profil proyeksi citra dokumen masukan (Zramdini dan Ingold, 1993). Misalnya terdapat sebuah citra biner S dengan banyaknya baris M dan banyaknya kolom N . Profil proyeksi vertikal (P_v) dari citra S , yaitu banyaknya piksel hitam yang tegak lurus sumbu y :

$$P_v[i] = \sum_{j=1}^M S[i, j] \quad (9)$$

Sedangkan profil proyeksi horisontal (P_h) dari citra S , yaitu adalah banyaknya piksel hitam yang tegak lurus sumbu x :

$$P_h[i] = \sum_{j=1}^N S[i, j] \quad (10)$$

Tahap berikutnya adalah tahap pengenalan karakter. Tahap ini akan diawali dengan proses ekstraksi ciri. Proses ekstraksi ciri adalah proses yang bertujuan untuk memperoleh karakteristik utama yang melekat pada suatu citra yang dapat membedakan antara satu citra dengan citra lainnya. Pada penelitian ini

pendekatan yang digunakan untuk menghasilkan ciri suatu citra karakter adalah pendekatan statistik. Secara garis besar salah cara yang dilakukan untuk mengekstrak ciri adalah sebagai berikut (Parker, 1997):

- Sebuah citra karakter dibagi menjadi 3×3 bagian sehingga menjadi 9 bagian yang disebut dengan unit.
- Setiap unit dalam keseluruhan citra akan dicari banyaknya piksel obyek.



Gambar 1. Ilustrasi perhitungan ciri karakter Jawa: tarung

Tahap kedua dalam pengenalan karakter dan sekaligus adalah tahap terakhir pengujian program adalah proses klasifikasi. Klasifikasi merupakan tahap pengenalan terhadap suatu obyek. Pada tahap ini suatu obyek akan dikelompokkan ke dalam suatu kelas tertentu berdasarkan ciri-cirinya. Pada penelitian ini, klasifikasi dilakukan dengan pendekatan alamiah dengan cara sebagai berikut:

- Nilai setiap unit pada lokasi yang bersesuaian dari karakter yang akan dikenali dibandingkan dengan nilai setiap unit karakter yang terdapat dalam basis data. Perbandingan ini dilakukan dengan menggunakan modifikasi dari jarak Euclidean. Apabila terdapat dua buah fitur karakter a, b dengan banyak unit sembilan, maka jarak $D_f(a, b)$ antara dua buah fitur tersebut dapat didefinisikan sebagai berikut:

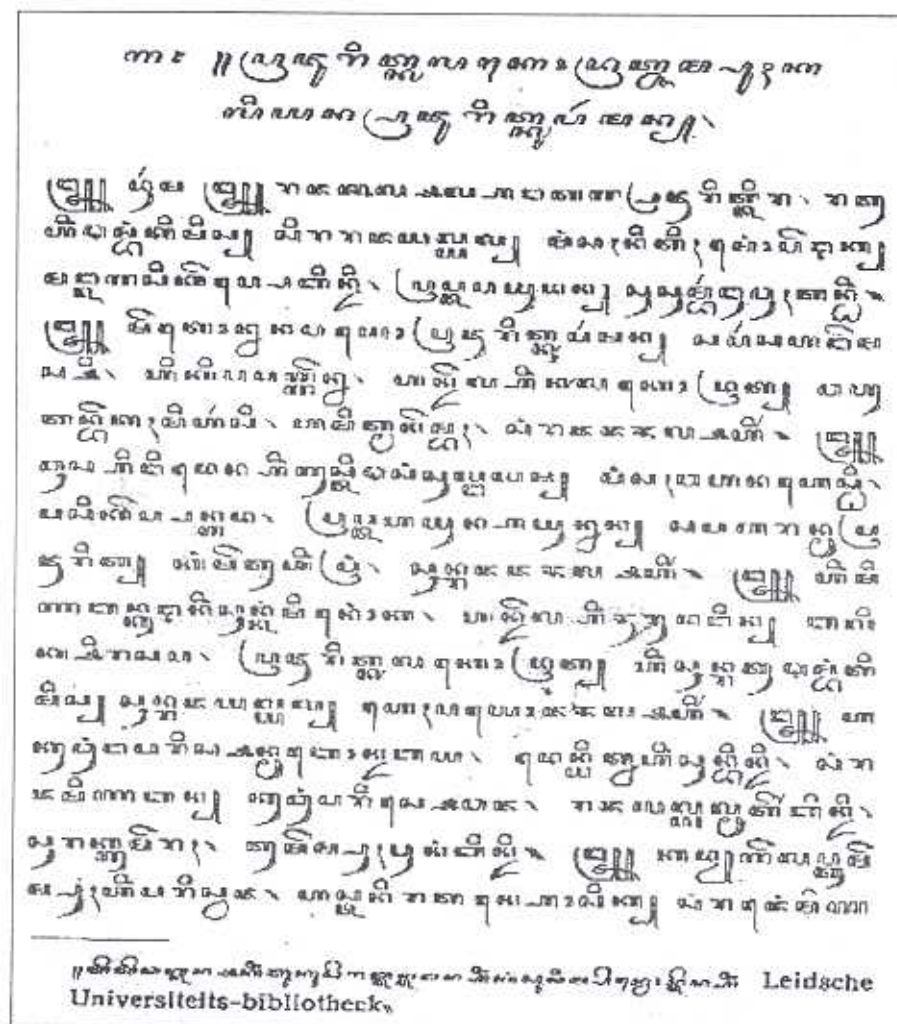
$$D_f(a, b) = \sum_{i=1}^3 \sum_{j=1}^3 |a_{ij} - b_{ij}| \quad (11)$$

- Dicari jarak yang paling minimum dari nilai-nilai hasil perbandingan pada langkah satu. Jarak yang paling minimum dan masih di bawah batas atas nilai jarak yang diperbolehkan akan dipilih sebagai karakter yang paling mendekati karakter yang akan dikenali tersebut.

HASIL DAN PEMBAHASAN

Dalam penelitian ini dilakukan pengujian program. Hasilnya kemudian dibandingkan dengan pembacaan secara manual. Pengujian program dimaksudkan untuk mengetahui kehandalan hasil keluaran program aplikasi MATLAB dibandingkan dengan hasil pembacaan secara manual. Penulis memilih hasil pembacaan dokumen secara manual oleh seorang sarjana S1 Sastra Jawa yang sekaligus juga adalah seorang pustakawan pada perpustakaan khusus untuk literatur Sastra Jawa.

Dari data dokumen Menak Sorangan I Halaman 3 seperti terlihat pada gambar 1, dengan *threshold* sebesar 20 (Widiarti, 2006) diperoleh hasil output pengenalan karakter seperti terlihat dalam tabel 2., yang sekaligus memperlihatkan bahwa tidak semua karakter dari citra dokumen dikenali secara benar. Hal tersebut dapat dilihat pada nama-nama latin yang ditandai dengan bayangan.



Gambar 1. Dokumen Menak Sorangan I Halaman 3

Tabel 2. Hasil Pengujian Program Untuk Dokumen Menak Sorangan I
Halaman 3

Baris	Karakter Dikenali
1	1 : adeg2 adeg2 pra ju ri t ka la taling ka tarung drat ca ma puh ka
2	li ya na pra ju rit ku par man PadaLingsa
3	PadaMadya Dur ma PadaMadya ra ja sa la sa la ha nga ta gal ga2 pra ju rit ti ra PadaLingsa ra tu
4	ing Tam bang ti mis si ra ra ja ya l ya l mang sah ni tih taling mang tarung pe thak
5	mang la ga si ke taling p pe bin di PadaLingsa prap ta pa yu dan su sum bar ngu suh tan ding PadaLingsi
6	PadaMadya me taling ta tarungn na na pa taling ya pra ju rit ku par man sa par sa ha nge ma
7	sa si PadaLingsa hi ki sa sag gen na PadaLingsa han de la ing Ka la taling ka PadaLingsa drat pa yu
8	tan be tah mi yar si PadaLingsa ya mit ma nem bah PadaLingsa sang ra ja ja Ba la sa hir PadaLingsi PadaMadya
9	wu sa i ngi taling da na ing gus ti Ta sang sul nga la m mang sah pa ha na taling has thi PadaLingsa
10	pa si kep pan ga da PadaLingsa prap ta ha yu na ha yun nan sa sa ha ran mu pra
11	jurit kang me tu ing pra PadaLingsa sun raja ja Ba lasa hir PadaLingsi PadaMadya ing mi
12	Na ban ku tha ning sun kang mi taling nang tarung ka PadaLingsa han de la ing Bu ru da ngin ba li
13	ka si ra sa sa PadaLingsa pra ju rit Ka la taling ka tarung drat ing sun ra tu Tam bang hi
14	mis sun ra ja ma l ya la1 la2 taling hah pa taling ya ja Ba la1 tarung sa hir PadaLingsi PadaMadya ha
15	ku dhung nga sa ri sa san mu taling ba tarungn ba ba ma tarung taling dan yit na ing sun bin di PadaLingsa sang ra
16	ja mi Na han ku dhung pa ri taling sa sa sa ja ,ra ja ya1 dal yal mu ter bin di PadaLingsa
17	su rak gu me rah PadaLingsa tu me ma1 ma2 puh pu nang bin di PadaLingsi PadaMadya kad ya ge la1 ma2p tu me
18	ma puh ing pa ris wa ja PadaLingsi has ta ni ra ta taling na ha tarung sik sa ra taling jang mi Na
19	adeg2 adeg2 te te dhak a na sa king bu ku se rat tan ta nga na ing kang taling mi ma pe talingn wa n te na ing l. na id
20	h h v t i b i o t k PadaLingsi

Dari hasil pengenalan karakter Tabel 2. diperoleh tabel data analisis *output* pengenalan citra karakter sebagai berikut :

Tabel 3. Analisis *Output* Citra Dokumen Menak Sorangan I Halaman 3

No	Jumlah Karakter Citra Dokumen					Tambahan Karakter	
	Karakter	Dikenali	Benar	Salah	Beda	Spasi	Karakter Lain
1	18	18	18	0	0		
2	11	12	11	0	1	1	
3	22	23	22	0	1		1
4	25	25	25	0	0	1	
5	26	27	25	1	1		1
6	24	24	23	1	0	1	
7	23	27	20	3	4	2	2
8	25	25	24	1	0		
9	26	26	25	1	0		
10	23	24	21	2	1		1
11	21	21	20	1	0		
12	24	30	24	0	6	6	
13	23	27	21	2	4	1	
14	23	25	20	3	2		2
15	26	28	22	4	2	2	
16	24	25	21	3	1		1
17	24	26	23	1	2		2
18	26	27	25	1	1		1
19	26	26	20	6	0		
20	25	26	11	14	1	1	
Total	465	492	421	44	27	1	11

Dari hasil analisis dalam Tabel 3. tampak bahwa, dari 465 karakter yang terdapat dalam citra dokumen, diperoleh 421 karakter yang dikenali secara benar. Sehingga dari hasil tersebut dapat dihitung prosentase derajat keberhasilan pengenalan citra yaitu sebesar $(421 / 465) \times 100\%$, yang akan bernilai 90.54%.

Cara-cara analisis pada hasil output dari pengenalan citra dokumen buku sastra Jawa Menak Sorangan 1 Halaman 3 tadi kemudian diterapkan pada analisis hasil output citra dokumen lain (lihat Tabel 1.). Sebagai hasil akhir diperoleh rangkuman hasil analisis *output* program sebagai berikut:

Tabel 4. Rangkuman Hasil Analisis Output Program

Nama Dokumen	Jumlah Karakter			Prosentase Keberhasilan
	Citra	Dikenali	Benar	
Menak Sorangan 1 Halaman 3	465	492	421	90.54
Menak Sorangan 1 Halaman 4	526	562	434	82.51
Panji Sekar Halaman 3	477	509	345	72.33
Panji Sekar Halaman 4	538	577	429	79.74
Rata-rata Prosentase Keberhasilan				81.28

Maka dari hasil analisis data diperoleh rata-rata keberhasilan proses pengenalan citra dokumen 81.28%. Nilai prosentase derajat keberhasilan ini menggambarkan bahwa pemilihan metode-metode dalam proses pengenalan citra dokumen yang digunakan dalam penelitian ini untuk mengenali citra dokumen dari buku sastra Jawa Modern Menak Sorangan 1 dan Panji Sekar relatif baik.

KESIMPULAN DAN SARAN

Dari penelitian yang telah dilakukan dapat ditarik kesimpulan bahwa untuk dokumen sastra Jawa modern Menak Sorangan 1 dan Panji Sekar halaman 3 dan 4, sistem telah dapat mengenali citra-citra karakter penyusun dokumen tersebut dengan rata-rata prosentase keberhasilan sebesar 81.28% dengan *threshold* jarak ciri citra karakter sebesar 20. Prosentase keberhasilan pengenalan citra dokumen ini selain tergantung pada variasi citra karakter yang tersimpan dalam basis data, juga sangat tergantung dari kualitas kertas, kualitas cetakan, dan kerumitan karakter cetakan dari dokumen yang akan diolah. Semakin baik kualitas kertas dan kualitas cetakan, semakin besar kemungkinan untuk menghasilkan olahan citra-citra karakter yang baik. Semakin rumit bentuk-bentuk karakter cetakan, semakin sulit proses pengolahan citra dokumennya.

Dari hasil penelitian ini tampak bahwa masih perlu dikaji lebih lanjut metode-metode lain baik untuk proses pengolahan tingkat piksel maupun proses pengenalan karakter sehingga diperoleh hasil pengolahan yang lebih optimal. Selain itu masih perlu dilakukan penelitian lebih lanjut untuk dokumen teks Jawa yang lain misalnya teks Jawa Kuno.

DAFTAR PUSTAKA

- Azar, D., 1997, "Hilditch's Algorithm for Skeletonization", <http://jeff.cs.mcgill.ca/~godfried/teaching/projects97/azar/skeleton.html#algorithm>.
- Brown, E.W., 2000, "Character Recognition by Feature Point Extraction", www.ccs.neu.edu/home/feberic/papers/charr ec.pdf.
- Fletcher, L.A., dan Kasturi, R., 1988, "A Robust Algorithm for Text String Separation from Mixed Text/Graphics Images", *IEEE Trans. Pattern Anal. Machine Intelli.*, Vol. 10, No.6, 910-918.
- Jain, A.K., 1989, *Fundamental of Digital Image Processing*, NJ, Prentice Hall Information and System Sciences Series.
- Kavallieratou, E., Fatokis, N., dan Kokkinakis, G., 2005, "Skew Angle Estimation In Document Processing Using Cohen's Class Distributions", <http://slt.wcl.ee.upatras.gr/papers/kavallieratou2.pdf>.
- O'Gorman, L., 1993, "The Document Spectrum for Page Layout Analysis", *IEEE Trans. Pattern Anal. Machine Intelli.*, Vol. 15, No. 11, 1162-1173.
- O'Gorman, L., dan Kasturi, R., 1997, *Executive Briefing: Document Image Analysis*, USA, IEEE Computer Society Press.
- Otsu, N., 2003, "A Threshold Selection Method from Gray-Level Histograms", www-users.itlabs.umn.edu/classes/Spring2003/csc8980/presentations/AThresholdSelectionMethodfromGray-Level.ppt.
- Pakubuwono IV., 1933, *Panji Sekar*, Batawisentrem, Bale Poestaka.

- Parker, J.R., 1997, *Algorithms For Image Processing And Computer Vision*, New York, Wiley Computer Publishing.
- Sauvola, J., Seppänen, T., Haapakoski, S., dan Pietikäinen, M., 1997, "Adaptive Document Binarization", www.mediateam.oulu.fi/publications/pdf/42.pdf.
- Srihari, S.N., Lam, S.W., Govindaraju, V., Srihari, R.K., dan Hull, J.J., 1986, *Document Image Understanding*, New York, CEDAR.
- Suprihatin, 2003, *Penerapan Finite Automata Dalam Mengalihaksarakan Tulisan Aksara Jawa ke Tulisan Huruf Latin. Tesis Program Studi Ilmu Komputer, Universitas Gadjah Mada, Yogyakarta*,
- Suryakusuma, S., 2003, "Kamus-Kamus Bahasa Jawa", http://www.tembi.org/perpus/2003_02_perpus03.htm
- Widiarti, A.R., 2006, *Pengenalan Citra Dokumen Sastra Jawa: Konsep dan Implementasinya. Tesis Program Studi Ilmu Komputer, Universitas Gadjah Mada, Yogyakarta*.
- Yasadipura I, R.Ng., 1936, *Menak Sorangan I*, Batawisentrem, Bale Poestaka.
- Zramdini, A., dan Ingold, R., 1993, "Optical font recognition from projection profiles", *Electronic Publishing*, Vol. 6, No.3, 249-260.